Paper 2

Abstract Title Page

Not included in page count.

Title: Power Calculations for Moderators in Multi-site Cluster Randomized Trials

Authors and Affiliations:

Jessaca Spybrook
Western Michigan University

Ben Kelcey
University of Cincinnati

Nianbo Dong
University of Missouri

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Cluster randomized trials (CRTs), or studies in which intact groups of individuals are randomly assigned to a condition, are becoming more common in evaluation studies of educational programs. A specific type of CRT in which clusters are randomly assigned to treatment within blocks or sites, known as multisite cluster randomized trials (MSCRTs), are the most frequent in the literature. In a recent review, Spybrook, Shi, & Kelcey (under review) noted that of 22 CRTs funded by the Institute of Education Sciences between 2011 and 2013, almost half (n=12) of the studies were MSCRTs in which students were nested in clusters and clusters were randomly assigned to condition within blocks or sites.

The primary question that often guides the design of MSCRTs is whether or not the program works. Hence the MSCRT is designed with the goal of being powered to detect the main effect of treatment. However, there are other important questions that one should consider during the design phase of the study. For example, for whom is the treatment effective? In what types of schools is the treatment effective? Hence the power to detect moderator effects may often enter the design considerations.

The power to detect moderator effects for MSCRTs at the student, cluster, or site level has a much smaller literature base than the power to detect the main effect of treatment for MSCRTs. In 2000, Raudenbush and Liu examined power calculations for multi-site studies in which individuals are randomly assigned within sites. Spybrook (2014) provided empirical estimates of minimum detectable effect sizes for moderator effects in a sample of MSCRTs funded by IES. Bloom & Spybrook (2013) calculated power to detect moderator effects as well as variability in treatment effects across sites for MSCRTs.

Purpose / Objective / Research Question / Focus of Study:

The purpose of this paper is to extend power calculations to moderator effects in MSCRTs. Specifically we consider 3-level MSCRTs in which individuals are nested within clusters and clusters are randomized within sites, and 4-level MSCRTs in which individuals are nested within sub-clusters, sub-clusters are nested within clusters and clusters are randomized within sites. We consider moderators at the individual, cluster, and site level and provide R code for the calculations. We demonstrate how power for main effects and moderator effects might both be considered in the design of a MSCRT.

Significance / Novelty of study:

This paper extends the power calculations for MSCRTs beyond the main effect of treatment to allow researchers to also consider the power for moderator effects in the design phase of the study. Currently, there is no user-friendly software for calculating power for moderator effects in MSCRTs. The R code provided in this study will help fill this gap in the resources available to researchers planning MSCRTs.

Statistical, Measurement, or Econometric Model:

We provide the models for a 3-level MSCRT. For purposes of the proposal, we focus on the main effect of treatment and moderator effects at the individual, cluster, and site level. The full paper includes all the models and the power calculations for the 4-level MSCRT.

Main effect of Treatment

The main effect of treatment is often the primary interest in a MSCRT. In this case, suppose we have students (level-1) nested within schools (level-2) and schools are randomly assigned to condition within districts (level-3). In the case of no moderator, the student level model is:

$$Y_{ijk} = \pi_{0jk} + e_{ijk} \qquad e_{ijk} \sim N(0, \sigma^2) \qquad [1]$$

for $i \in \{1,2,...,n\}$ persons per cluster, $j \in \{1,2,...,J\}$ clusters and $k \in \{1,2,...,K\}$ sites, where π_{0jk} is the mean for cluster j in site k; e_{ijk} is the error associated with each student; and σ^2 is the within-cluster variance.

The level-2 model, or cluster-level model, is:

$$\pi_{0jk} = \beta_{00k} + \beta_{01k} W_{jk} + r_{0jk} \qquad r_{0jk} \sim N(0, \tau_\pi) \qquad [2]$$

where β_{00k} is the mean for site k; β_{01k} is the treatment effect at site k; W_{jk} is a treatment contrast indicator, ½ for treatment and -½ for the control; r_{0jk} is the random effect associated with each cluster; and τ_π is the variance between clusters within sites.

The level-3 model, or site-level model, is:

$$\begin{aligned} \beta_{00k} &= \gamma_{000} + u_{00k} && \text{var}\ (u_{00k}) \sim \tau_{\beta_{00}} \\ \beta_{01k} &= \gamma_{010} + u_{01k} && \text{var}\ (u_{01k}) \sim \tau_{\beta_{11}} \qquad \text{cov}(u_{00k}, u_{01k}) = \tau_{\beta_{01}} \end{aligned} \qquad [3]$$

where γ_{000} is the grand mean; γ_{010} is the average treatment effect ("main effect of treatment"); u_{00k} is the random effect associated with each site mean; u_{01k} is the random effect associated with each site treatment effect; $\tau_{\beta_{00}}$ is the variance between site means; $\tau_{\beta_{11}}$ is the variance between sites on the treatment effect; and $\tau_{\beta_{01}}$ is the covariance between site-specific means and site-specific treatment effects. Note that we allow the treatment effect to vary randomly across sites, however, we could also treat this as a fixed effect.

The estimate of the treatment effect and the variance of the estimated treatment effect are:

$$\hat{\gamma}_{010} = \bar{Y}_E - \bar{Y}_C$$

$$Var\left(\hat{\gamma}_{010}\right) = \left[\tau_{\beta_{11}} + 4\left(\tau_{\pi} + \sigma^2 / n\right)/ J\right]/ K \qquad [4]$$

The power for the test for the main effect of treatment for the 3-level MSCRT, $H_0 : \gamma_{010} = 0$, follows the same logic as the power for the main effect of treatment for the 2-level CRT (Raudenbush, 1997). The *F*statistic in this case though is a ratio $MS_{treatment}$ to the $MS_{treatmentbycluster}$. The ratio of expected mean squares is equivalent to 1+ λ, where the noncentrality parameter is defined as:

$$\lambda = \frac{\gamma_{010}^2}{\left[\tau_{\beta_{11}} + 4\left(\tau_{\pi} + \sigma^2 / n\right)/ J\right]/ K} \quad \text{or} \quad \lambda = \frac{\delta^2}{\left[\sigma_{\delta}^2 + 4\left(\rho + (1-\rho)/ n\right)/ J\right]/ K} \qquad [5]$$

where $$\delta = \frac{\gamma_{010}}{\sqrt{\tau_{\pi} + \sigma^2}}, \rho = \frac{\tau_{\pi}}{\tau_{\pi} + \sigma^2}, \sigma_{\delta}^2 = \frac{\tau_{\beta_{11}}}{\tau_{\pi} + \sigma^2}$$

We standardize the parameters by the sum of the within site variance. As the noncentrality parameter increases, the power of the test increases. Note that the model above could easily be extended to include covariates to increase the precision of the estimate.

Moderator Effects

Given the space limitations in the proposal, we do not provide the models for each of the possible moderator effects: students, school, and district level. Instead, we provide the non-centrality parameter for each case. The parameters are standardized as shown in equation (5) for all cases. In each case, we assume a binary moderator, i.e. gender, 75 percent or greater free/reduced lunch in school vs. less than 75 percent free/reduced lunch in school. For student and school moderators, we treat them as fixed effects in higher level of the model. Note that we assume one moderator at the specific level in each case. Table 1 provides the noncentrality parameter for each of the three cases.

Insert Table 1 about here

Findings / Results:

Table 1 suggests that the influence of the sample sizes at each level is not the same across the moderator effects and the main effect of treatment. That is, for the individual level moderator effect, the number of individuals per cluster has more influence in terms of increasing power than it does for the main effect or cluster or site moderator effect. The power to detect the cluster level moderator effect is more greatly influenced by the number of individuals per cluster and number of clusters per site than the power for the main effect or site moderator effect. Finally, the power for the site level moderator effect is influenced primarily by the number of sites, similar to the main effect, however the within site variance is four times that of the main effect of

treatment, suggesting larger numbers of sites are needed to detect site moderator effects than the main effect.

Conclusions:

It is critical to consider not only the main effect of treatment in the design of MSCRTs but also important moderator effects such as whether the treatment if more effective for boys than girls, or in schools with a high percent of kids with free/reduced lunch versus schools with a low percent of kids with free/reduced lunch. Three sample sizes contribute to the power in all cases, the number of individuals per cluster, the number of clusters per site, and the total number of sites. The relative influence of each sample size varies depending on the effect of interest: main effect, individual level moderator effect, cluster level moderator effect, or site level moderator effect. For example, increasing the number of individuals per cluster yields greater increases in power for the individual level moderator than in would for the main effect of treatment. Similarly increasing the number of sites is critical if the goal is to power for the main effect of treatment and the site level moderator. The important effects of interest, i.e. main effects and specific moderator effect, should be considered in the design phase of the study in order to try to maximize the likelihood that a study is powered to answer the primary questions.

Appendix A. References

Bloom, H.S., & Spybrook, J. (2013). *Statistical power/precision for multi-site trials: aka The power hour or precision decisions*. Presentation for the the William T. Grant Foundation workshop on Learning from Variation in Program Effects.

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods, 2*(2), 173-185.

Raudenbush, S. W., & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods, 5*(2), 199-213.

Spybrook, J. (2014). Detecting intervention effects across context: An Examination of the Power of Cluster Randomized Trials. *Journal of Experimental Education*, *82(3), 334-357*.

Spybrook, J., Shi,R., & Kelcey, B. (Under Review). Progress in the past decade: An examination of the precision of cluster randomized trials funded by the U.S. Institute of Education Sciences.

Appendix B. Tables and Figures

Table 1. The standardized noncentrality parameter for each type of moderator effect.

Individual Level Moderator*	Cluster Level Moderator**	Site Level Moderator***
$\lambda = \frac{K\delta^2}{\omega + 4(\kappa + 4(1-\rho)/n)/J}$	$\lambda = \frac{K\delta^2}{\nu + 16(\rho + (1-\rho)/n)/J}$	$\lambda = \frac{K\delta^2}{4\sigma^2_{\vert s} + 16\left(\rho + (1-\rho)/n\right)/J}$

*For individual level moderator, ω is standardized variance of moderator effect across sites, κ is standardized variance of moderator within sites, δ is the standardized individual moderator effect, all other parameters are same as in case of main effect of treatment.
**For cluster level moderator, ν is the standardized variance of the moderator effect across sites, δ is the standardized individual moderator effect, all other parameters are same as in case of main effect of treatment.
***For site level moderator, $\sigma^2_{\delta|s}$ is the standardized residual variance of the moderator effect across, δ is the standardized individual moderator effect, all other parameters are same as in case of main effect of treatment.